SCIENTIFIC REPORTS

OPEN

Dissecting the expression relationships between RNA-binding proteins and their cognate targets in eukaryotic post-transcriptional regulatory networks

Received: 25 February 2016
Accepted: 21 April 2016
Published: 10 May 2016

Sneha Nishtala[1], Yaseswini Neelamraju[1] & Sarath Chandra Janga[1,2,3]

RNA-binding proteins (RBPs) are pivotal in orchestrating several steps in the metabolism of RNA in eukaryotes thereby controlling an extensive network of RBP-RNA interactions. Here, we employed CLIP (cross-linking immunoprecipitation)-seq datasets for 60 human RBPs and RIP-ChIP (RNP immunoprecipitation-microarray) data for 69 yeast RBPs to construct a network of genome-wide RBP-target RNA interactions for each RBP. We show in humans that majority (~78%) of the RBPs are strongly associated with their target transcripts at transcript level while ~95% of the studied RBPs were also found to be strongly associated with expression levels of target transcripts when protein expression levels of RBPs were employed. At transcript level, RBP - RNA interaction data for the yeast genome, exhibited a strong association for 63% of the RBPs, confirming the association to be conserved across large phylogenetic distances. Analysis to uncover the features contributing to these associations revealed the number of target transcripts and length of the selected protein-coding transcript of an RBP at the transcript level while intensity of the CLIP signal, number of RNA-Binding domains, location of the binding site on the transcript, to be significant at the protein level. Our analysis will contribute to improved modelling and prediction of post-transcriptional networks.

Progress in proteomics together with other omics technologies have now convincingly shown the existence of an additional and perhaps more important gene regulatory layer in cellular networks, which acts in concert with other layers of regulation to control gene expression and translation in a highly coordinated complex system defined as post-transcriptional regulatory network. For instance, in one of the large-scale omics studies comparing transcriptome and proteome levels it was shown that ~30% of the variance in protein abundance in yeast cannot be explained by mRNA expression levels[1]. Comparison of the dynamic transcriptome and proteome profiles in yeast also revealed the presence of several classes of post-transcriptionally regulated proteins, accounting for more than 40% of the proteome[2]. In another study, a comparison of functional clusters inferred from transcriptome and translatome data in yeast revealed the presence of three groups of proteins: transcriptionally co-regulated proteins cluster together in transcriptome as well as translatome data and represent metabolic processes; post-transcriptionally co-regulated proteins cluster together only in translatome data and consist of RNA-binding, ribosomal and protein synthesis proteins; and dually co-regulated proteins have intermediate co-clustering characteristics and hence are likely regulated at both levels[3]. Increasing number of studies now suggest that the lack of mRNA-protein correlation in eukaryotic cells can be explained due to the post-transcriptional control mediated by several regulatory RNAs with the major protein players being

[1]Department of Bio Health Informatics, School of Informatics and Computing, Indiana University Purdue University, 719 Indiana Ave Ste 319, Walker Plaza Building, Indianapolis, Indiana 46202, USA. [2]Centre for Computational Biology and Bioinformatics, Indiana University School of Medicine, 5021 Health Information and Translational Sciences (HITS), 410 West 10th Street, Indianapolis, Indiana, 46202, USA. [3]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Medical Research and Library Building, 975 West Walnut Street, Indianapolis, Indiana, 46202, USA. Correspondence and requests for materials should be addressed to S.C.J. (email: scjanga@iupui.edu)

RNA-binding proteins (RBPs)[4,5]. Recent studies also show that RNA-binding proteins (RBPs) which play a crucial role in the post-transcriptional regulation of gene expression[4–6] themselves exhibit distinct expression dynamics in post-transcriptional regulatory networks[7] and tend to bind functionally related mRNAs with most mRNAs bound by multiple RBPs, resulting in a complex network of post-transcriptional regulatory interactions[8,9].

In prokaryotes, functionally related genes are often organized into operons to facilitate co-expression and to reduce expression fluctuation among the resulting protein products[10]. Indeed, coordinated regulation of functionally related genes by reducing their expression variation in a cell is important for the survival of organisms with limited resources and has been observed in eukaryotes as well[11–14]. Hence, it was proposed that posttranscriptional regulons in eukaryotes may play an equivalent role to operon structures in prokaryotes in coordinating the expression of their target genes during posttranscriptional regulation[15–17]. According to the RNA regulon theory, trans-acting factors like RBPs combinatorially regulate multiple mRNAs to achieve functionally coherent translation in the face of stochastic gene transcription[14]. This posttranscriptional regulation of genes is important in splicing, transport, localization, translational control, stability and degradation of RNAs[9,14,18]. These various phases of RNA metabolism are regulated when the RBPs bind to the RNAs to form RNP complexes[4,9]. Therefore, the fate of RNA is dictated by the interaction of RNAs with the RBPs within the RNP complexes[6,19]. Various RBP mediated events have been well documented using expression profiles which are specific to tissues and conserved across different species[20–23]. With large amount of transcriptomic and proteomic data and a multitude of RBPs being identified, it has become possible to test if RBPs can direct the expression of their target transcripts using various flavors of RNA interactome datasets for RBPs in yeast and other model systems[24–26]. In particular, crosslinking immunoprecipitation (CLIP)-seq technology[27] has proven to be a potent tool in the study and understanding of the transcriptomic *in-vivo* binding sites of RBPs at the single nucleotide level[28]. While experimental studies indicate that the function of RBPs on gene expression is complicated and sometimes can exhibit opposite trends depending on the growth condition, it is unclear whether RBPs can modulate expression levels of their target transcripts in humans and if there is an association between them in post-transcriptional regulatory networks[29–32]. Although a recent study suggests that RBPs are co-regulated with their target genes and plays an important role in coordinating their expression variation in yeast[25], it is not clear how prevalent is this phenomenon and what factors contribute to such associations. In humans, numerous diseases have been linked to the defects in RBP function[33–35]. With many examples of RBPs being identified, it becomes feasible to test whether these post-transcriptional regulons can coordinate the expression of their target transcripts on a genome-wide level[25] at least in model systems such as yeast and human with large-scale interactome data for multiple RBPs[8,36,37], offering a unique opportunity to examine the regulatory relationships between RBPs and their target mRNAs[3,7,25,26].

In this study, we map the CLIP binding sites of 60 RBPs on to the human genome to construct a RBP – RNA network. Further, we examined the correlation of each RBP's expression at both transcript and protein levels with the target RNAs to see how these correlation patterns change. We then analyzed different factors impacting the change in expression patterns through a comprehensive two level analysis using different modelling techniques namely multivariate regression modelling, stepwise linear regression and Elastic net. We observed a higher level of association between the protein expressions of RBPs with their target transcripts compared to transcript level expression. Our results indicate that RBPs at both proteomic and transcriptomic levels play an important role in coordinating expression changes of the target RNAs and this can be explained by various factors governing the functions of these RBPs.

Results

Overview of the analysis. As shown in Fig. 1, we downloaded CLIP data for 60 human RBPs from the CLIPdb database and used hg19 annotations from Ensembl database to build post-transcriptional regulatory networks linking RBPs to their target transcripts for each RBP (see Materials and Methods and Table 1). As discussed in Materials and Methods, we mapped the binding sites of each RBP to 300 bps upstream and downstream flanking regions of each exon and considered its corresponding transcript to be a target of the RBP if the binding sites map on these regions of the exons (Fig. 1). This allowed us to construct a genome-wide network of RBP-RNA interactions linking RBPs to their target transcripts in the human genome as summarized in Table 1. The target transcript annotations were compared with the quantified transcript level expression data across 16 human tissues from the Human Body Map (HBM) project and were divided into three groups of transcripts for each RBP as described in Materials and Methods using ad hoc python scripts (Fig. 1). For each RBP, both the transcript level of one protein coding transcript and its protein expression data from Human Protein Atlas (HPA) were independently used for correlation analysis with its target transcripts as described in Materials and Methods. Similar approach was employed for analyzing the associations between RBPs and their target transcripts in the constructed post-transcriptional regulatory network of the yeast genome (see Materials and Methods). To understand the different factors influencing the observed correlation patterns at both the transcript and protein levels in the human genome, we undertook a comprehensive modelling approach using three different feature selection/reduction methods at two different levels – RBP centric and transcript centric level by considering the different factors listed in Tables 2 and 3. A discussion of the selected features is presented in the Materials and Methods and the respective results section.

Majority of the RBPs exhibit significant association with their target transcripts at the transcript level. To understand how RBPs are associated with their targets, we correlated the expression of one of the protein coding transcripts (with the highest mean expression level across all the tissues) of each RBP with the expression of the target transcripts across 16 human tissues and compared it with control set of transcripts (all transcripts which do not belong to the class of target transcripts). To address the issue of the size of the control set, we also randomly sampled the control set of transcripts by sampling 100 times to extract each time

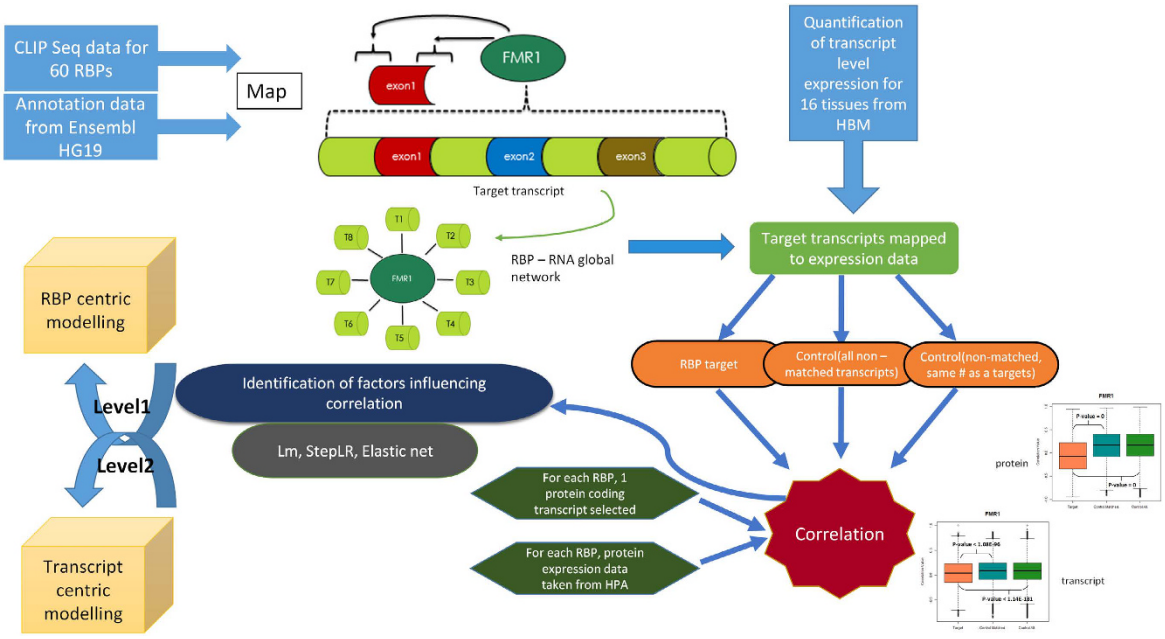

Figure 1. Flowchart summarizing the major steps involved in the construction and analysis of the human post-transcriptional regulatory network controlled by RBPs, employed in this study. Data required for the analysis was downloaded from CLIP DB[28] and Ensembl[51]. The binding sites of each RBP were mapped to target transcripts such that, if the binding site of the RBP falls within the 300 bps flanking regions and 300 bps upstream regions of at least one of its annotated exonic start or end coordinates, its corresponding transcript would be considered a target transcript of the RBP. A global network was created for each RBP. The target transcripts were mapped on to the RNA-seq expression data from the Human Body Map (HBM)[52–54] and three categories of transcripts were constructed based on whether a transcript group is targeted by an RBP or not. One protein coding transcript for each RBP with highest mean expression level across all 16 tissues was chosen and spearman correlation was calculated with transcripts from each of the three categories of transcripts. Similarly, correlations between protein expression levels of RBPs and their target/non-target transcripts expression levels from corresponding matched RNA-seq samples were calculated using protein expression data downloaded from Human Protein Atlas (HPA)[55]. Different patterns in associations between the RBPs and the three categories of transcripts were identified and classified. To explain the observed associations, three different feature selection/reduction methods were employed at two different levels, namely – RBP centric and transcript centric level.

the same number of randomly selected transcripts as the number of target transcripts, which is referred to as the control-matched set. As a result of computing the spearman correlations between a RBP transcript and its target as well as non-target (control) transcripts, RBPs were divided into three classes based on comparing the distribution of correlation coefficients for targets *versus* control associations; 1) Significantly Congruent (SC) : RBPs would belong to this class if the distribution of RBP – target correlation coefficients have their median correlation coefficient significantly higher than that seen in the control set of transcripts (Wilcoxon test, $p < 0.05$) 2) Significantly incongruent (SIC) : RBPs would belong to this class if the distribution of RBP – target correlation coefficients have their median correlation coefficient significantly lower than that seen in the control set of transcripts (Wilcoxon test, $p < 0.05$) 3) No Significant change (NSC) : If no significant change in the median correlation coefficient is observed between the targets and the control sets, those RBPs would belong to this class. RBPs grouped into these three classes are supported by robust set of p–values as illustrated in Supplementary Fig. 1. Figure 2 shows the six most significant SC and SIC RBPs at the transcript level. PTBP1 with PTB domain[38,39], known to control pathways related to translational control and splicing and CSTF2T for mRNA – splicing[40] are among the significant SC RBPs. FMR1 known to be important for translation control and documented to be implicated in several neurological disorders[41] as well as LIN28A known for cardiac progenitor differentiation and translational control[42] were found to be among the significant SIC RBPs. Supplementary Fig. 1 shows boxplots comparing the correlation coefficients for target *versus* control transcripts for all the 60 RBPs organized into SC (36 RBPs), SIC (11 RBPs) and NSC (13 RBPs) classes. Overall, we found that 78.33% of the RBPs comprised of SC and SIC classes, exhibited significant association with their targets at the transcript level, at a p-value threshold of 0.05.

Significant fraction of the yeast RBPs also exhibit an association with their target transcripts revealing the conservation of expression coupling. To understand, whether our observation of finding RBP – target expression correlations to be significantly non-random, is generic and conserved across organisms, we analyzed the RBP – RNA network of the yeast genome using the same workflow (see Materials and Methods). Since humans and yeast are evolutionary distant, we hypothesized that the yeast genome would be an

RBP	# targets	Source of CLIP-Seq data (References)
AGO1	17,206	67,68
AGO2	64,425	67–76
AGO3	5,849	67
AGO4	1,517	67
ALKBH5	1,685	77
ATXN2	6,715	78
C17ORF85	2,199	77
CAPRIN1	7,506	77
CPSF1	9,132	79
CPSF2	1,813	79
CPSF3	2,832	79
CPSF4	3,486	79
CPSF6	38,973	79
CPSF7	45,556	79
CSTF2	36,829	79,80
CSTF2T	32,893	79
DGCR8	14,587	81
EIF4A3	35,729	82
ELAVL1	55,432	69,83–85
EWSR1	6,702	86,87
EZH2	116	88
FBL	4,019	89
FIP1L1	34,643	79
FMR1	10,732	90
FUS	2,150	87,91,92
FXR1	2,940	90
FXR2	9,230	90
HNRNPA1	9,587	93
HNRNPA2B1	2,013	93
HNRNPC	18,806	94,95
HNRNPD	6,121	96
HNRNPF	3,461	93
HNRNPH	3,608	97
HNRNPM	3,051	93
HNRNPU	10,318	93,98
IGF2BP1	15,328	67
IGF2BP2	10,668	67
IGF2BP3	10,022	67
LIN28A	16,860	42,99
LIN28B	21,277	99,100
MOV10	7,523	101
NOP56	2,007	89
NOP58	7,879	89
NUDT21	40,541	79
PTBP1	25,385	102,103
PTBP2	16,956	102,103
PUM2	1,938	67
QKI	1,358	67
RTCB	6,376	77
SRRM4	10,026	103
TAF15	3,197	87,104
TARDBP	12,138	105
TIA1	11,977	106
TIAL1	23,854	106
TNRC6A	828	67
Continued		

RBP	# targets	Source of CLIP-Seq data (References)
TNRC6B	643	67
TNRC6C	859	67
WDR33	1,626	107
YTHDF2	15,474	108
ZC3H7B	13,294	77

Table 1. List of human RBPs, their number of target transcripts and source of CLIP data. Table shows a list of all the 60 RBPs employed in the analysis, the number of transcripts targeted by each of them and references to the studies which provide the CLIP-Seq data[28]. This list was generated by mapping the binding sites of each RBP with exonic coordinates and obtaining the corresponding transcripts of the mapped exons.

Variable	Feature name	Description
Response	Median correlation coefficient of target transcripts	This was calculated by taking the median of all the correlation coefficients of each RBP with its target transcripts.
Predictor	Number of target transcripts	The number of target transcripts of each RBP was obtained by the mapping the co-ordinates of the binding sites onto the annotated transcripts as described in materials and methods.
	Median CLIP Signal	CLIP peaks in the bed format obtained from CLIPdb[28] come with a P- value signifying the intensity of the CLIP binding for each binding site. The median P - value of the binding sites which were mapped to the targets, for each RBP, was calculated.
	Number of RNA binding domains	Number of RNA binding domains for each RBP was obtained from a previous study describing the compendium of human RBPs [5].
	Number of protein-protein interactions	Human protein - protein interaction network was constructed using data from BIOGRID[48]. Then, for each RBP, its number of interacting partners was computed.
	Number of protein coding transcripts	The number of protein coding transcripts documented for each RBP was obtained from Ensembl[51].
	Number of annotated transcripts	Total number of transcripts (protein coding, processed transcript, etc.,) documented for each RBP was also obtained from Ensembl[51].
	Length of the selected protein coding transcript	The length of the selected protein coding transcript used for computing correlation of each RBP with target transcripts and control transcripts was also obtained from Ensembl[51].
	Median distance of binding site from transcript	The closest distance of the start of the binding site from either ends on the transcript was calculated for each target transcript. The median value of this distance was then calculated for each RBP.
	mRNA - protein correlation	The protein expression of each RBP was correlated with the mRNA expression across nine tissues with both RNA and protein expression data.

Table 2. Different features employed to study their contribution towards observed correlation between RBPs and their post-transcriptional targets, in the RBP centric modelling. For each feature details about how it was computed is also listed.

ideal model to show the generality of our observations across different species. The network used in this study[43] comprises of 69 RBPs corresponding to 24,932 RBP – RNA interactions. On performing a similar analysis to that described for the human RBPs, we found comparable results in the yeast genome. In particular, the correlation patterns revealed that 63.08% (41/65) of the RBPs display an association (Wilcoxon test, $p < 0.05$) (Supplementary Fig. 2). Among the RBPs which exhibited a significantly higher/lower correlation coefficient compared to control transcripts, 58.54% (24) could be classified as SIC and 41.46% (17) as SC RBPs. Figure 3 shows the six most significant SC and SIC RBPs. YJL080C, commonly known as - SCP160 is important for mRNA metabolism in yeast[44] and interestingly, in our analysis, it is shown to be one of the most significantly associated RBPs with its target mRNAs. Similarly, YIR034C or LYS1 is important for mRNA binding in yeast[45,46] and is one among the highly associated SIC RBPs. Supplementary Fig. 2 shows all the 65 RBPs organized into SC, SIC and NSC classes.

Most RBPs exhibit significant association with their target transcripts at the protein level. While we found that RBPs show good degree of associations with their targets in both the human and yeast genomes when the transcript levels of RBPs are employed, protein expression levels of RBPs in matched tissues or experimental conditions is rather limited. However, recent genome-wide protein levels for multiple human tissues resulting from the Human Proteome Map facilitate addressing this question, albeit using limited number of samples (see Materials and Methods). After identifying and mapping equivalent RNA-seq and proteomic samples, we correlated the protein expression data across nine tissues for each RBP with the corresponding target as well as control transcripts' expression levels from the RNA-seq dataset and organized the RBPs in to three classes – SC, SIC and NSC. Figure 4 shows six most significant SC and SIC RBPs at the protein level. Several members of CPSF and HNRNP family exhibited significant correlation with their targets often in different directions. Supplementary Fig. 3 shows all the 58 RBPs organized into the three different classes – SC (11 RBPs), SIC (44 RBPs) and NSC (3 RBPs). Note that two of the RBPs, RTCB and SRRM4 had no expression levels documented in the protein expression dataset and hence were not included in this analysis. Overall, we found that ~95% of the RBPs exhibited significant association with their target transcripts at the protein level (Wilcoxon test, $p < 0.05$). These results support the notion that the protein expression levels of most RBPs are strongly correlated with their target transcripts expression levels. Indeed, this association is far stronger than that observed at the transcript level of an RBP.

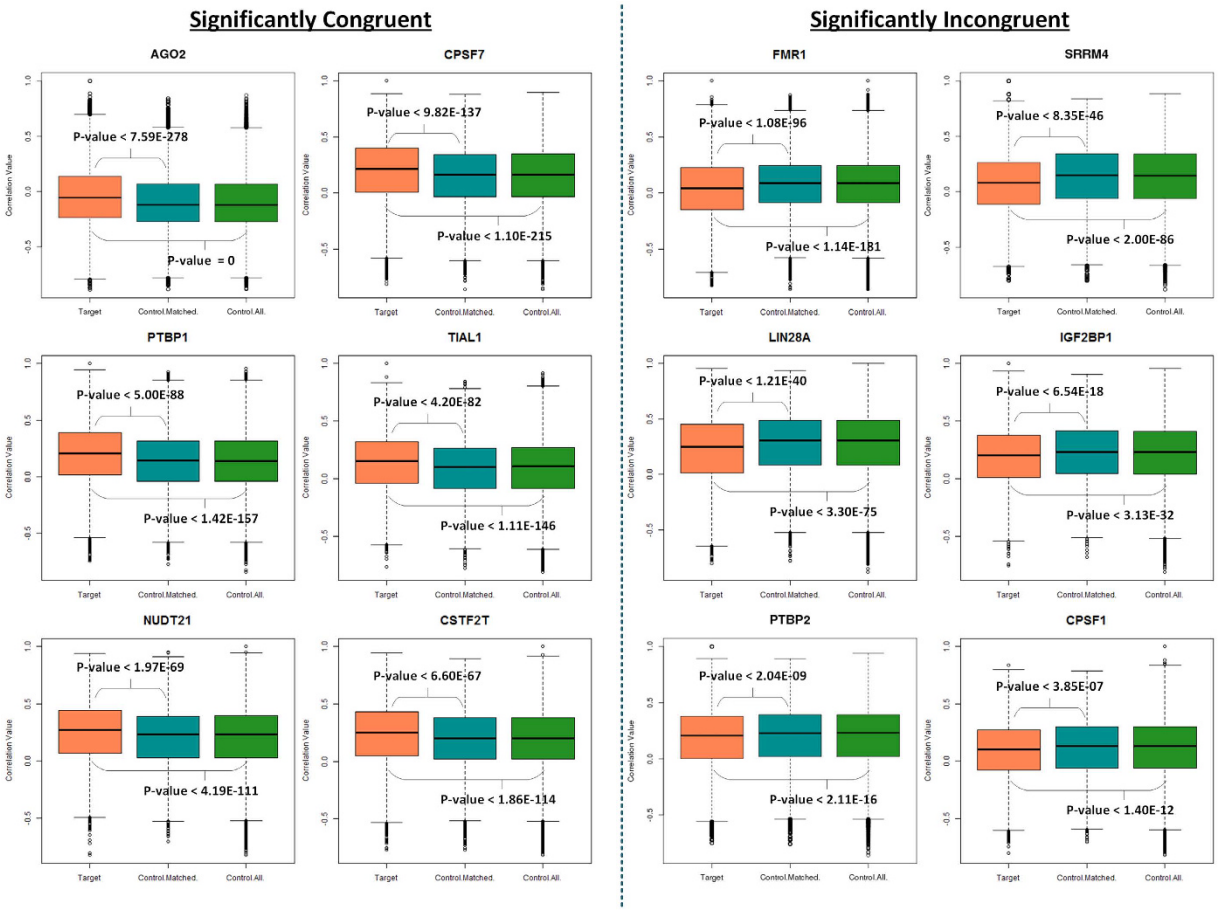

Figure 2. Selected set of six human RBPs each belonging to the significantly congruent and incongruent categories, when only transcriptome data was used for computing the correlations between RBPs and their target/non-target transcripts. Boxplots showing the distribution of correlation coefficients between RBPs and transcripts belonging to the three categories, red: protein coding transcript of RBP correlated with its target transcripts, blue: protein coding transcript of RBP correlated with the same number of random non-targeted transcripts as the number of target transcripts in the post-transcriptional regulatory network of an RBP, green: protein coding transcript of RBP correlated with all the non-targeted transcripts.

Only a small fraction of the RBPs show similar patterns of associations with their targets at both the protein and transcript levels. To further understand and dissect our findings on RBP – target associations, we performed a comparative analysis of the outcomes for various human RBPs to see how these correlation patterns change within the transcript or protein levels and from the transcript to the protein levels of RBPs. Figure 5A summarizes the results of our analysis by showing the percentage of RBPs showing associations at the transcript and protein levels, number of RBPs falling into each of the three classes - SC, SIC and NSC, while Fig. 5B shows a heatmap showing the significance values (−log(p-value)) of the RBP – target associations compared to 100 matched control sets used in the previous sections. We find that at the transcript level 60% of the RBPs fall into the SC category while at the protein level, 18.97% fall into this category. Likewise, 18.33% RBPs fall into the SIC category at the transcript level while 75.86% RBPs fall into this category at the protein level.

Further, we observed that 12 RBPs exhibited similar trends at both the transcript and protein levels with six of them belonging to SC category and six belonging to the SIC category. To further understand the behavior of these 12 RBPs (sync RBPs) and how they are different compared to others (non-sync RBPs), we analyzed different network centrality measures using igraph[47] package in R. This was achieved by constructing a protein interaction network for RBPs using two different sources - Biogrid databas[48] and String database[49] separately, to obtain an unbiased understanding of the differences in the network centrality measures irrespective of the dataset used. We found that closeness centrality for sync RBPs is higher than non-sync RBPs using interaction networks from both Biogrid and String databases (Wilcoxon test, $p < 3.55E\text{-}08$ and $p < 1.80E\text{-}10$ respectively) indicating that sync RBPs have shorter average path lengths to other proteins in the protein interaction network and hence must be well connected to other proteins (Supplementary Table 1). Using the interaction network from String database, we also found betweenness centrality to be different between the sync and non-sync groups (Wilcoxon test, $p < 0.006$). Therefore, we postulate that sync RBPs are likely functionally active and remain the same at the protein level too.

Variable	Feature name	Description
Response	Correlation coefficient of RBP with target transcript	For each RBP, the correlation coefficient of the selected protein coding transcript (or protein expression at protein level) with each target transcript was selected as the response variable.
Predictor	CLIP Signal	CLIP peaks in the bed format obtained from CLIPdb[28] come with a P- value signifying the intensity of the CLIP binding for each binding site. The median P - value of the binding sites which were mapped on to the target, for each RBP, was calculated.
	Distance of binding site from 5′ end	The distance of the start of the binding site from the 5′ end of each target transcript was calculated.
	Distance of binding site from 3′ end	The distance of the start of the binding site from the 3′ end of each target transcript was calculated.
	Transcript Length	The length of each target transcript was obtained from Ensembl[51].
	Transcript Type	The biotype of each target transcript was also obtained from Ensembl[51].

Table 3. Different features employed to study their contribution towards observed correlation between RBPs and their post-transcriptional targets, in the transcript centric modelling. For each feature details about how it was computed is also listed.

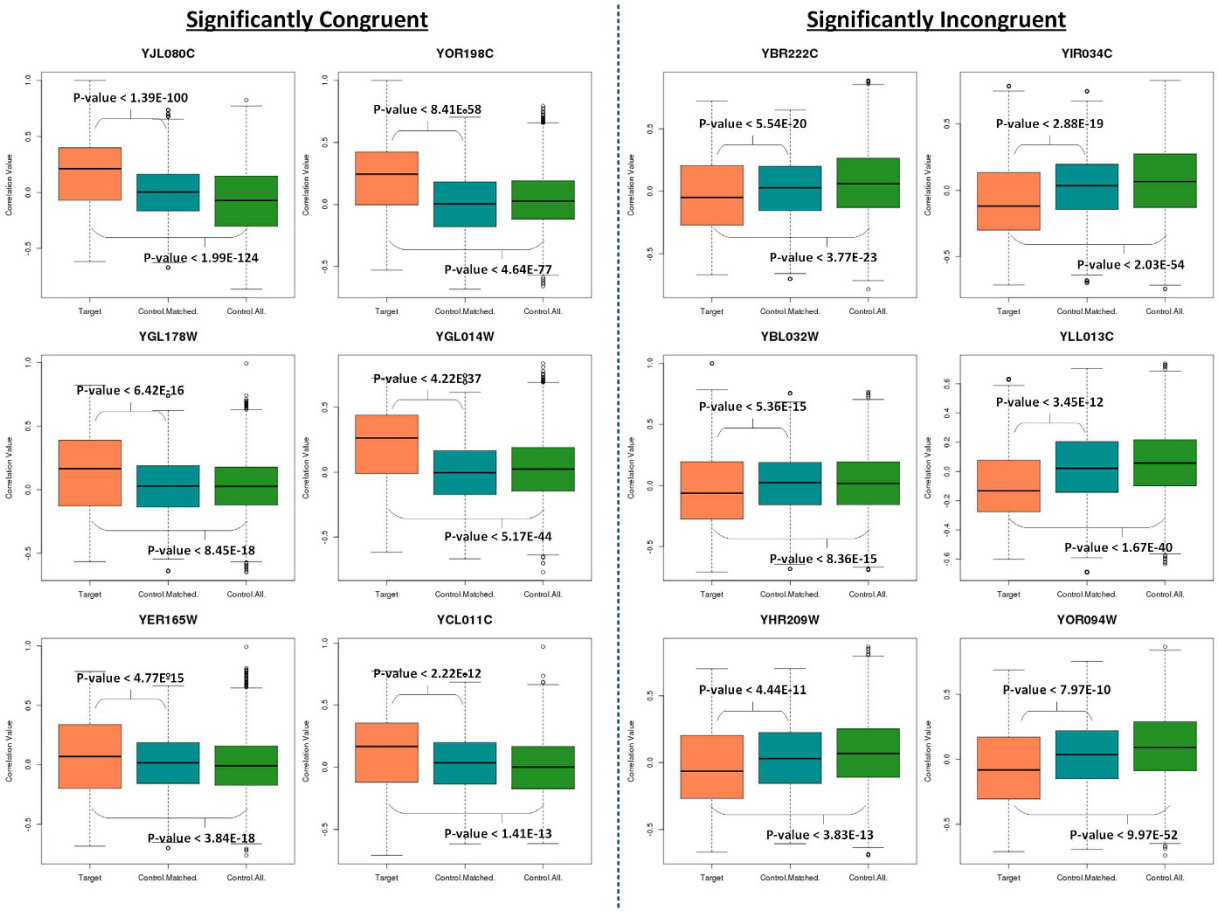

Figure 3. Selected set of six yeast RBPs each belonging to the significantly congruent and incongruent categories from the yeast post-transcriptional regulatory network of RBPs[43]. Boxplots showing the distribution of correlation coefficients between RBPs and transcripts belonging to the three categories, red: transcript expression of RBP correlated with its target transcripts, blue: transcript expression of RBP correlated with the same number of random non-targeted transcripts as the number of target transcripts in the post-transcriptional regulatory network of an RBP, green: transcript expression of RBP correlated with all the non-targeted transcripts.

Different set of features influence the correlation observed at the transcript and protein levels. As listed in Table 2 and discussed in Materials and Methods, nine features were selected which we hypothesized to contribute to the observed correlation patterns between RBPs and their transcripts. We selected these features for RBP centric modelling because each of these features are likely contributing to the function or dynamics of an RBP, its influence on the target transcript or can be attributed to the strength of its binding signal on the target transcript in either a direct or indirect mode and hence would therefore be explanatory of the observed

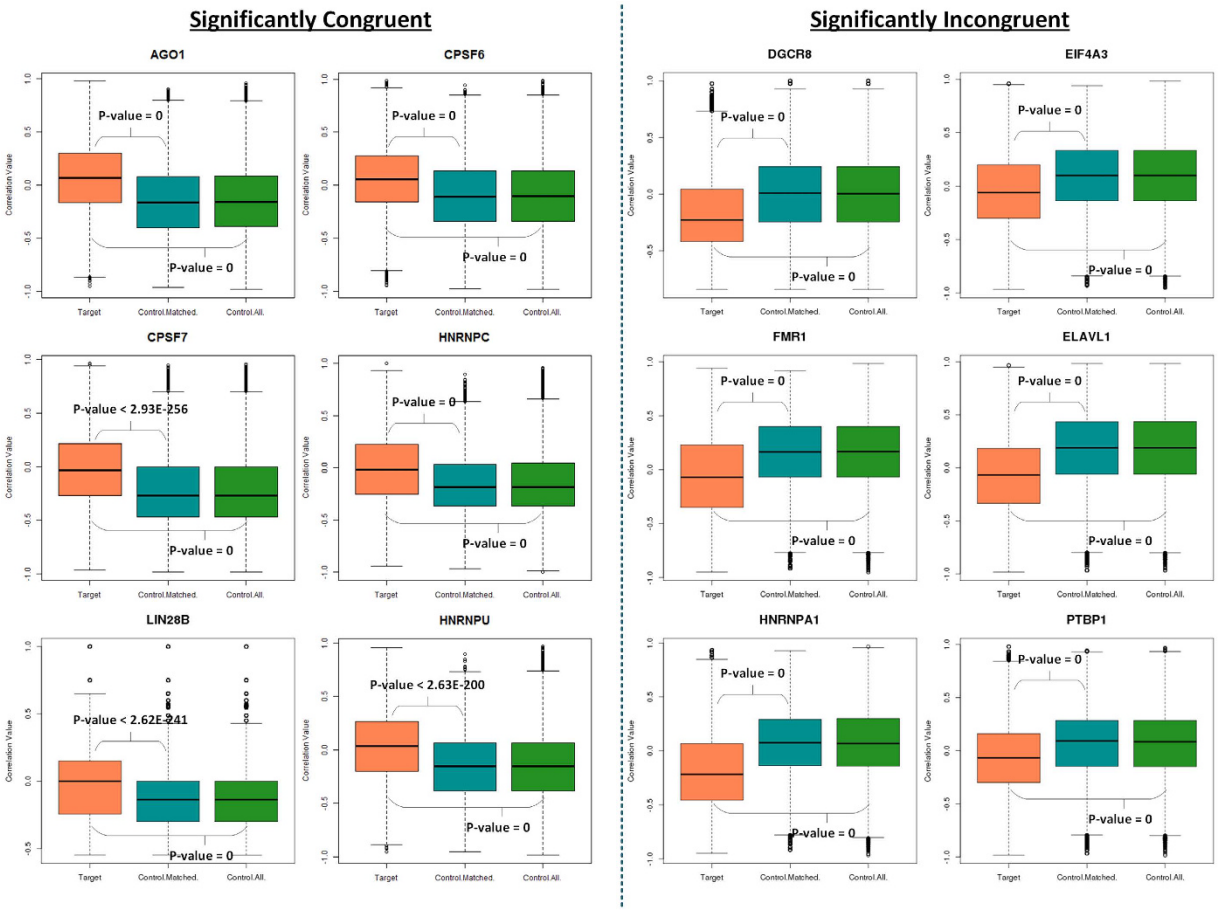

Figure 4. Selected set of six human RBPs each belonging to the significantly congruent and incongruent categories, when protein expression levels of RBPs and transcript levels of the target/non-targets was used for computing the correlations. Boxplots showing the distribution of correlation coefficients between RBPs and transcripts belonging to the three categories, red: protein expression level of an RBP correlated with its target transcripts, blue: protein expression level of an RBP correlated with the same number of random non-targeted transcripts as the number of target transcripts in the post-transcriptional regulatory network of an RBP, green: protein expression level of an RBP correlated with all the non-targeted transcripts.

trend. We used three different feature selection approaches, multivariate regression, step wise linear regression and elastic net to identify a reproducible and robust set of important features. Figure 6a shows the significance (−log(p-value) for all the features tested, at the transcript and protein level plotted as a heatmap. We found that at the transcript level, the number of target transcripts, number of protein coding transcripts and the length of the selected protein coding transcript were the most important features while at the protein level, median clip signal, number of RNA binding domains and median distance of the binding site on the transcript as important features. Similar results were obtained using the elastic net framework. Supplementary Fig. 4 displays the important features obtained using this method.

Type of the transcript and distance of the binding site from either side of the transcript prove to be important features at the transcript centric level. As listed in Table 3, five features were selected which we hypothesized to contribute to the observed correlation patterns between RBPs and their transcripts at the transcript level modelling (see Materials and Methods). Briefly, in transcript centric modelling, the response variable is the correlation coefficient between each RBP and its target transcript essentially enumerating all possible RBP-transcript pairs. As earlier, we used three different methods similar to the procedure described for the RBP centric modeling to uncover the robust set of features. Figure 6b,c show the significance (−log(p-value)) of the different features for the various RBPs using multivariate regression modeling. A similar figure without clustering is available as Supplementary Fig. 5, for easier reference. We found that transcript type is a significant feature for 66.67% and 63.80% of the RBPs at the transcript and protein levels respectively. There were a total of 56 transcript types that were available from Ensembl BioMart[50] into which our target transcripts were classified. We therefore tried to understand which transcript type is more contributing to the response variable by inspecting the median correlation coefficient of each transcript type for each RBP showing transcript type as an important feature. We observed that interchangeably, protein coding transcript and processed transcript followed by miRNA

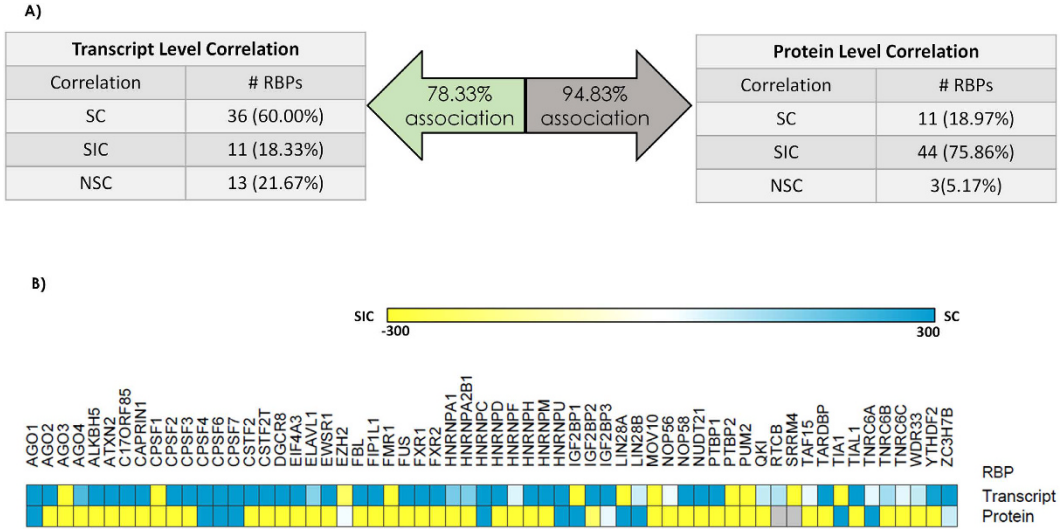

Figure 5. Summary of the various correlation patterns observed at the transcriptomic and proteome levels for human RBPs. (**A**) Distribution of RBPs into SC, SIC and NSC categories when the RNA and protein expression levels of the RBP respectively, are considered for computing the correlations. (**B**) Heatmap showing the significance (−log(p-value)) of the observed correlation compared to that seen in random non-targeted transcripts for various RBPs when the RNA and protein expression levels of an RBP are considered. P-values are computed using the Wilcoxon test comparing the distributions of correlation coefficients between targeted and 100 sets of non-targeted transcripts for each RBP. In the heatmap, blue color represents the SC RBPs and the yellow color represents the SIC RBPs.

and lncRNA were the important transcript types impacting the expression correlation. We also found that the distance of the binding site of the RBP from 3′ or 5′ end of the transcript were important factors for many RBPs.

Discussion

In summary, we find that RBPs exhibit significant co-expression patterns with their target RNAs although the extent and direction of co-expression can vary between RBPs and among members of the same RBP family. They show strong association with their targets at both protein and transcript levels, however a higher level of association was observed at the protein level. Most of the RBPs show different level of association at the protein and transcript level with only 20% of them showing similar trends at both the levels. Intensity of the clip signal, number of RNA binding domains and location of the binding site on the transcript prove to be important features which can explain the association observed at the protein level while number of target transcripts, number of protein coding transcripts for the selected RBP and length of selected protein coding transcript explain the associations seen at the transcript level. On further dissecting the analysis, at the transcript centric level, we observe that the type of target transcript and the distance of the binding site from the 5′ or 3′ end of the transcript are the important factors. We also found contrasting trends i. e, same RBP can be an SC at the transcript level while being classified as an SIC RBP at the protein level - classified based on the expression associations of the RBPs with their target transcripts at the transcript and protein levels. It is interesting to note that different features were found to be significant at these two levels possibly suggesting the rational for the observed differences in directionality of the associations.

In this study, we present genome-scale evidence that majority of the RBPs are correlated in expression levels with their post-transcriptionally controlled target transcripts in both the human and yeast genomes. To our knowledge this is the first study to report such an association in post-transcriptional regulatory networks and strengthens our understanding of the relationship between RBPs and their cognate targets. Our observations suggest that in disease conditions, expression associations between RBPs and target transcripts are likely altered. Hence, prognostic RBPs can be identified by comparing the extent and number of associations in healthy versus disease expression cohorts, enabling a means of rapidly profiling for RBP biomarkers in developmental diseases, cancer and other complex disorders[4,5,33]. In conclusion, our study provides a deeper insight into the behavior of various RBPs in the context of post-transcriptional regulatory networks. Thus, providing a roadmap for the identification of different post-transcriptional regulatory patterns thereby enabling rational design of experiments pertaining to protein – RNA associations.

Materials and Methods

Datasets employed for human RBP binding sites as well as tissue-specific RNA and protein expression levels. To obtain a comprehensive understanding of the RBP-RNA interaction networks on a genome- wide scale and to study the characteristics of binding sites of RBPs on their target RNAs, we downloaded Crosslinking Immunoprecipitation followed by high-throughput sequencing (CLIP-Seq) data from CLIPdb database[28] for 60 RBPs in humans. Although there is data for 63 RBPs in CLIPdb, we limited our analysis to those

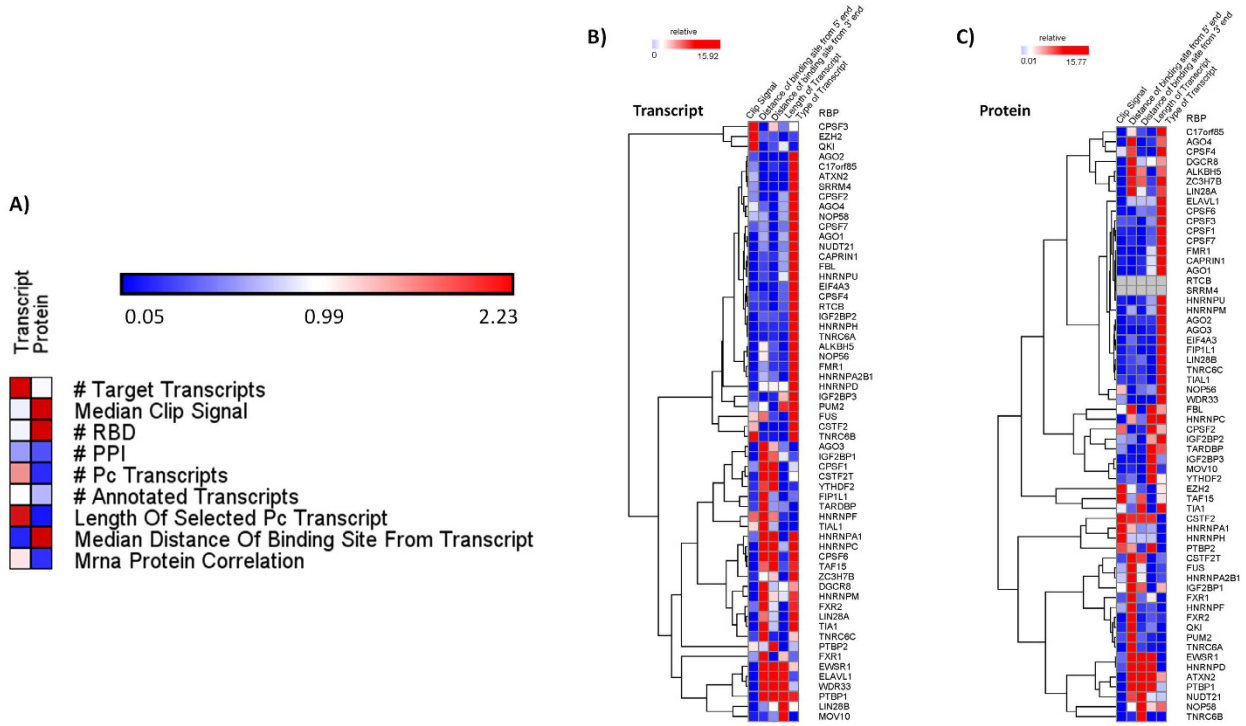

Figure 6. Heatmaps showing the significance of various features influencing correlation at the RBP and transcript centric levels. Heatmaps show the significance values (−log(p-value)) obtained by performing multivariate regression modelling to predict features influencing correlation at the (**A**) RBP level. (**B**,**C**) Transcript level when RNA and protein levels of RBPs were considered. Significance values for various features considered in this analysis are clustered hierarchically. Similar results were obtained using stepwise linear regression and elastic net regression modelling.

RBPs with high quality CLIP-seq data, limiting the number to 60 RBPs. We obtained the complete set of 217,426 annotated transcripts for the human genome from Ensembl using Biomart[50,51]. We mapped the binding sites onto the Ensembl HG19 version 79[51] of the human genome to find all the target RNAs for each RBP. RNA-seq data for 16 human tissues from Illumina's Human Body Map (HBM) 2.0[52–54] was downloaded from ArrayExpress database (http://www.ebi.ac.uk/arrayexpress) under the accession number E-MTAB-513. To study the correlation of protein expression levels of RBPs with RNA levels of RBP target transcripts, we downloaded mass spectrometry based proteomic data for over 17,000 human proteins across 30 tissues/cell lines (17 adult, 7 fetal tissues and 6 hematopoietic cells) from Human Proteome Atlas (HPA)[55].

Constructing RBP-RNA regulatory network. Several studies have shown that RBPs bind 200–300 nucleotides around the observed splice sites, which generally possess the identifiable sequence features[56,57]. We therefore considered a transcript to be a target of a RBP, if and only if the binding sites of the RBP fall within the 300 bps flanking regions or 300 bps downstream regions of at least one of its annotated exonic start or end co-ordinates. Based on this criterion, if at least one exon is mapped with a RBP's binding site, the corresponding transcript is considered its target transcript. This allowed us to build a RBP – target transcript network for each RBP which was used in the downstream analysis (see Fig. 1). The union of unique number of transcripts targeted by each of the 60 RBPs is 121,131 transcripts. The number of target transcripts for each RBP based on the built regulatory network is listed in Table 1.

Correlation analysis to study the association between each RBP and its target transcripts. Transcript level expression was quantified for the downloaded RNA-seq data using Sailfish v0.6.3[58]. TPM (transcripts per million) values were considered for the quantification of expression levels of all the Ensembl annotated transcripts in the human genome. The target transcripts which were mapped as described above for each of the 60 RBPs, were then matched with the genome-wide transcript levels across tissues and were divided into 3 categories for each RBP – 1) RBP target transcripts 2) RBP control matched – defined as the set of randomly selected transcripts in the expression compendium equal in number as the number of target transcripts and 3) RBP control all – defined as the set of all the transcripts which were not annotated to be targeted by a RBP based on CLIP-seq data.

To identify a representative protein coding transcript encoding for an RBP among all the annotated transcripts in the RNA-seq data, a protein coding transcript with the highest mean expression level across all the 16 tissues was chosen. Spearman correlation was calculated between the transcript level expression of each RBP and every target as well as non-target transcript's expression levels across the 16 tissues to generate correlation coefficients for the three different categories of transcripts namely RBP target transcripts, RBP control matched and RBP

control all, defined above. Similarly, for the protein expression data upon mapping the tissues from this dataset with the tissues available in the RNA-seq dataset, we found 9 common tissues and hence spearman correlations were computed between the protein expression level of each RBP and its target as well as non-target transcript's expression level from RNA-seq data which resulted in three categories of correlation coefficients for each RBP (see Fig. 1). Boxplots were plotted to represent the differences in the extents of correlation among the three categories of transcripts for each RBP and corresponding pairwise Wilcoxon test p-values computed using R to understand the significance of the observed patterns.

Identification of factors contributing to the observed association between RBPs and their target transcripts. To understand what factors and the extent to which they might be contributing to the observed correlation patterns at the transcript and protein levels of RBPs, we employed multivariate modelling at two different levels, referred to as the RBP centric level and the transcript centric level in this study. We employed three different feature selection/reduction methods to identify the robust set of contributing features, namely – the lm function, the step LR function and the ElasticNet[59] package in R. The lm function is an inbuilt function in R with a typical model in the form response ~ terms and is used to fit linear models and carry out regression and analysis of covariance and variance. Contrary to the lm method where all the features are included in the analysis, step function is an automated procedure where at the end of each step, variables with the most insignificant p – values are dropped and the procedure stops when the remaining features have a p – value significantly defined by a threshold value alpha. Ridge regression (L2 regularization term)[60] uses all input features to fit a model, while LASSO (L1 regularization term)[61] tries to find the most optimal fit. Elastic Net is a regularized regression modelling method which combines the above two methods and optimizes the bias and variance discrepancies between lasso and ridge.

At the RBP centric level, our goal was to understand and identify the general features which can likely explain the observed association between RBPs and their targets. These included nine features namely number of target transcripts controlled by an RBP, median CLIP signal of a RBP, number of RNA-binding domains in a RBP, number of documented protein interactions of a RBP, number of protein coding transcripts encoded by a RBP, total number of annotated transcripts by the gene encoding for RBP, length of the selected protein coding transcript, median of all the distances between the binding site of RBP and to the closest end of the transcript and correlation between mRNA-protein levels of an RBP, which could play a role in influencing the median correlation coefficient of all the target transcripts for each RBP. At the transcript centric level, our goal was to uncover the contribution of the transcript specific features such as CLIP signal on the target transcript, distance of binding site with respect to the 5′ or 3′ end of the transcript, length as well as type of the transcript on the correlation coefficient of each target transcript for each RBP. A detailed description of each of these features is listed in Tables 2 and 3 for RBP centric and transcript centric models respectively. RNA-binding domain annotations for RBPs were obtained from a previous study[5] and the number of protein – protein interactions for each RBP was calculated by constructing a protein – protein interaction network using interaction data from the BIOGRID database[48].

Post-transcriptional regulatory network of RBP – RNA interactions in yeast and analysis of correlation patterns. We hypothesized that the observed correlation patterns of RBPs with their target transcripts is conserved across species. To test this hypothesis, we used the post-transcriptional regulatory network of 69 RBPs and 24,932 RBP-RNA interactions in the yeast genome[43]. We downloaded RNA-seq data from a previous study generated under 18 different environmental conditions in yeast, with each condition having two biological replicates[62]. In particular, the data available at http://downloads.yeastgenome.org/published_datasets/Waern_2013_PMID_23390610/ for *S. cerevisiae* strain S288C reference genome sequence version R64-1-1[63] were downloaded from Saccharomyces Genome Database[64]. Raw data was quality filtered, aligned using Tophat[65] and expression levels of transcripts quantified using Cufflinks[66]. Similar analysis as implemented for the human genome was executed to identify the association between RBPs and their target/non-target RNAs. Boxplots were plotted using R. We limited our analysis to only 65 RBPs as opposed to 69 RBPs since for four of the yeast RBPs, only 1 target was detected with negligible expression levels.

References

1. Lu, P., Vogel, C., Wang, R., Yao, X. & Marcotte, E. M. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* **25,** 117–124, doi: 10.1038/nbt1270 (2007).
2. Lee, M. V. *et al.* A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. *Mol Syst Biol* **7,** 514, doi: 10.1038/msb.2011.48 (2011).
3. Joshi, A., Van de Peer, Y. & Michoel, T. Structural and functional organization of RNA regulons in the post-transcriptional regulatory network of yeast. *Nucleic Acids Res* **39,** 9108–9117, doi: 10.1093/nar/gkr661 (2011).
4. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nature reviews. Genetics* **15,** 829–845, doi: 10.1038/nrg3813 (2014).
5. Neelamraju, Y., Hashemikhabir, S. & Janga, S. C. The human RBPome: From genes and proteins to human disease. *Journal of proteomics*, doi: 10.1016/j.jprot.2015.04.031 (2015).
6. Glisovic, T., Bachorik, J. L., Yong, J. & Dreyfuss, G. RNA-binding proteins and post-transcriptional gene regulation. *FEBS letters* **582,** 1977–1986, doi: 10.1016/j.febslet.2008.03.004 (2008).
7. Mittal, N., Roy, N., Babu, M. M. & Janga, S. C. Dissecting the expression dynamics of RNA-binding proteins in posttranscriptional regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 20300–20305, doi: 10.1073/pnas.0906940106 (2009).
8. Hogan, D. J., Riordan, D. P., Gerber, A. P., Herschlag, D. & Brown, P. O. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS biology* **6,** e255, doi: 10.1371/journal.pbio.0060255 (2008).
9. Janga, S. C. From specific to global analysis of posttranscriptional regulation in eukaryotes: posttranscriptional regulatory networks. *Briefings in functional genomics* **11,** 505–521, doi: 10.1093/bfgp/els046 (2012).

10. Zaslaver, A., Mayo, A., Ronen, M. & Alon, U. Optimal gene partition into operons correlates with gene functional order. *Phys Biol* **3,** 183–189, doi: 10.1088/1478-3975/3/3/003 (2006).
11. Hornstein, E. & Shomron, N. Canalization of development by microRNAs. *Nat Genet* **38** Suppl S20–24, doi: 10.1038/ng1803 (2006).
12. Cui, Q., Yu, Z., Purisima, E. O. & Wang, E. MicroRNA regulation and interspecific variation of gene expression. *Trends Genet* **23,** 372–375, doi: 10.1016/j.tig.2007.04.003 (2007).
13. Li, X., Cassidy, J. J., Reinke, C. A., Fischboeck, S. & Carthew, R. W. A microRNA imparts robustness against environmental fluctuation during development. *Cell* **137,** 273–282, doi: 10.1016/j.cell.2009.01.058 (2009).
14. Keene, J. D. RNA regulons: coordination of post-transcriptional events. *Nature reviews. Genetics* **8,** 533–543, doi: 10.1038/nrg2111 (2007).
15. Keene, J. D. & Tenenbaum, S. A. Eukaryotic mRNPs may represent posttranscriptional operons. *Mol Cell* **9,** 1161–1167 (2002).
16. Halbeisen, R. E. & Gerber, A. P. Stress-dependent coordination of transcriptome and translatome in yeast. *PLoS biology* **7,** e1000105, doi: 10.1371/journal.pbio.1000105 (2009).
17. Wu, C. I., Shen, Y. & Tang, T. Evolution under canalization and the dual roles of microRNAs: a hypothesis. *Genome Res* **19,** 734–743, doi: 10.1101/gr.084640.108 (2009).
18. Halbeisen, R. E., Galgano, A., Scherrer, T. & Gerber, A. P. Post-transcriptional gene regulation: from genome-wide studies to principles. *Cell Mol Life Sci* **65,** 798–813, doi: 10.1007/s00018-007-7447-6 (2008).
19. Castello, A. *et al.* Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149,** 1393–1406, doi: 10.1016/j.cell.2012.04.031 (2012).
20. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22,** 1760–1774, doi: 10.1101/gr.135350.111 (2012).
21. Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* **338,** 1593–1599, doi: 10.1126/science.1228186 (2012).
22. Brawand, D. *et al.* The evolution of gene expression levels in mammalian organs. *Nature* **478,** 343–348, doi: 10.1038/nature10532 (2011).
23. Chan, E. T. *et al.* Conservation of core gene expression in vertebrate tissues. *Journal of biology* **8,** 33, doi: 10.1186/jbiol130 (2009).
24. Cirillo, D. *et al.* Constitutive patterns of gene expression regulated by RNA-binding proteins. *Genome biology* **15,** R13, doi: 10.1186/gb-2014-15-1-r13 (2014).
25. Jiang, H., Xu, L., Wang, Z., Keene, J. & Gu, Z. Coordinating expression of RNA binding proteins with their mRNA targets. *Scientific reports* **4,** 7175, doi: 10.1038/srep07175 (2014).
26. Pancaldi, V. & Bahler, J. *In silico* characterization and prediction of global protein-mRNA interactions in yeast. *Nucleic Acids Res* **39,** 5826–5836, doi: 10.1093/nar/gkr160 (2011).
27. Konig, J., Zarnack, K., Luscombe, N. M. & Ule, J. Protein-RNA interactions: new genomic technologies and perspectives. *Nature reviews. Genetics* **13,** 77–83, doi: 10.1038/nrg3141 nrg3141 [pii] (2011).
28. Yang, Y. C. *et al.* CLIPdb: a CLIP-seq database for protein-RNA interactions. *BMC genomics* **16,** 51, doi: 10.1186/s12864-015-1273-2 (2015).
29. Foat, B. C., Houshmandi, S. S., Olivas, W. M. & Bussemaker, H. J. Profiling condition-specific, genome-wide regulation of mRNA stability in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 17675–17680, doi: 10.1073/pnas.0503803102 (2005).
30. Saint-Georges, Y. *et al.* Yeast mitochondrial biogenesis: a role for the PUF RNA-binding protein Puf3p in mRNA localization. *PLoS One* **3,** e2293, doi: 10.1371/journal.pone.0002293 (2008).
31. Klein, M. E., Younts, T. J., Castillo, P. E. & Jordan, B. A. RNA-binding protein Sam68 controls synapse number and local beta-actin mRNA metabolism in dendrites. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 3125–3130, doi: 10.1073/pnas.1209811110 (2013).
32. Zhao, W. *et al.* Massively parallel functional annotation of 3′ untranslated regions. *Nat Biotechnol* **32,** 387–391, doi: 10.1038/nbt.2851 (2014).
33. Kechavarzi, B. & Janga, S. C. Dissecting the expression landscape of RNA-binding proteins in human cancers. *Genome biology* **15,** R14, doi: 10.1186/gb-2014-15-1-r14 (2014).
34. Lukong, K. E., Chang, K. W., Khandjian, E. W. & Richard, S. RNA-binding proteins in human genetic disease. *Trends in Genetics* **24,** 416–425, doi: S0168-9525(08)00173-X [pii] 10.1016/j.tig.2008.05.004 (2008).
35. Cooper, T. A., Wan, L. & Dreyfuss, G. RNA and disease. *Cell* **136,** 777–793, doi: 10.1016/j.cell.2009.02.011 (2009).
36. Gerber, A. P., Herschlag, D. & Brown, P. O. Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS biology* **2,** E79, doi: 10.1371/journal.pbio.0020079 (2004).
37. Scherrer, T., Mittal, N., Janga, S. C. & Gerber, A. P. A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. *PLoS One* **5,** e15499, doi: 10.1371/journal.pone.0015499 (2010).
38. Licatalosi, D. D. *et al.* Ptbp2 represses adult-specific splicing to regulate the generation of neuronal precursors in the embryonic brain. *Genes & development* **26,** 1626–1642, doi: 10.1101/gad.191338.112 (2012).
39. Margolis, B., Borg, J. P., Straight, S. & Meyer, D. The function of PTB domain proteins. *Kidney international* **56,** 1230–1237, doi: 10.1046/j.1523-1755.1999.00700.x (1999).
40. Romeo, V., Griesbach, E. & Schumperli, D. CstF64: cell cycle regulation and functional role in 3′ end processing of replication-dependent histone mRNAs. *Molecular and cellular biology* **34,** 4272–4284, doi: 10.1128/MCB.00791-14 (2014).
41. Verkerk, A. J. *et al.* Alternative splicing in the fragile X gene FMR1. *Human molecular genetics* **2,** 399–404 (1993).
42. Wilbert, M. L. *et al.* LIN28 binds messenger RNAs at GGAGA motifs and regulates splicing factor abundance. *Mol Cell* **48,** 195–206, doi: 10.1016/j.molcel.2012.08.004 (2012).
43. Mittal, N., Scherrer, T., Gerber, A. P. & Janga, S. C. Interplay between posttranscriptional and posttranslational interactions of RNA-binding proteins. *Journal of molecular biology* **409,** 466–479, doi: 10.1016/j.jmb.2011.03.064 (2011).
44. Lang, B. D. & Fridovich-Keil, J. L. Scp160p, a multiple KH-domain protein, is a component of mRNP complexes in yeast. *Nucleic Acids Res* **28,** 1576–1584 (2000).
45. Tsvetanova, N. G., Klass, D. M., Salzman, J. & Brown, P. O. Proteome-wide search reveals unexpected RNA-binding proteins in Saccharomyces cerevisiae. *PLoS One* **5,** doi: 10.1371/journal.pone.0012671 (2010).
46. Borell, C. W., Urrestarazu, L. A. & Bhattacharjee, J. K. Two unlinked lysine genes (LYS9 and LYS14) are required for the synthesis of saccharopine reductase in Saccharomyces cerevisiae. *Journal of bacteriology* **159,** 429–432 (1984).
47. Nepusz, G. C. a. T. The igraph software package for complex network research. *InterJournal* Complex Systems, 1695 (2006).
48. Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **34,** D535–539, doi: 10.1093/nar/gkj109 (2006).
49. Szklarczyk, D. *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* **39,** D561–568, doi: 10.1093/nar/gkq973 (2011).
50. Kinsella, R. J. *et al.* Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database the journal of biological databases and curation* **2011,** bar030, doi: 10.1093/database/bar030 (2011).
51. Flicek, P. *et al.* Ensembl 2014. *Nucleic Acids Res* **42,** D749–755, doi: 10.1093/nar/gkt1196 (2014).

52. Asmann, Y. W. *et al.* Detection of redundant fusion transcripts as biomarkers or disease-specific therapeutic targets in breast cancer. *Cancer research* **72,** 1921–1928, doi: 10.1158/0008-5472.CAN-11-3142 (2012).
53. Barbosa-Morais, N. L. *et al.* The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338,** 1587–1593, doi: 10.1126/science.1230612 (2012).
54. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* **22,** 1775–1789, doi: 10.1101/gr.132159.111 (2012).
55. Kim, M. S. *et al.* A draft map of the human proteome. *Nature* **509,** 575–581, doi: 10.1038/nature13302 (2014).
56. Fu, X. D. & Ares, M. Jr. Context-dependent control of alternative splicing by RNA-binding proteins. *Nature reviews. Genetics* **15,** 689–701, doi: 10.1038/nrg3778 (2014).
57. Jangi, M., Boutz, P. L., Paul, P. & Sharp, P. A. Rbfox2 controls autoregulation in RNA-binding protein networks. *Genes & development* **28,** 637–651, doi: 10.1101/gad.235770.113 (2014).
58. Patro, R., Mount, S. M. & Kingsford, C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat Biotechnol* **32,** 462–464, doi: 10.1038/nbt.2862 (2014).
59. Hastie, H. Z. a. elasticnet: Elastic-Net for Sparse Estimation and Sparse PCA. (2012).
60. Hoerl A. E., K. R. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* **12,** (1970).
61. R, T. Regression shrinkage and selection via the Lasso. *J Roy Stat Soc B Met* (1996).
62. Waern, K. & Snyder, M. Extensive transcript diversity and novel upstream open reading frame regulation in yeast. *G3* **3,** 343–352, doi: 10.1534/g3.112.003640 (2013).
63. Engel, S. R. *et al.* The reference genome sequence of Saccharomyces cerevisiae: then and now. *G3* **4,** 389–398, doi: 10.1534/g3.113.008995 (2014).
64. Cherry, J. M. *et al.* Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* **40,** D700–705, doi: 10.1093/nar/gkr1029 (2012).
65. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14,** R36, doi: 10.1186/gb-2013-14-4-r36 (2013).
66. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28,** 511–515, doi: 10.1038/nbt.1621 (2010).
67. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141,** 129–141, doi: 10.1016/j.cell.2010.03.009 (2010).
68. Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495,** 333–338, doi: 10.1038/nature11928 (2013).
69. Kishore, S. *et al.* A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat Methods* **8,** 559–564, doi: 10.1038/nmeth.1608 (2011).
70. Gottwein, E. *et al.* Viral microRNA targetome of KSHV-infected primary effusion lymphoma cell lines. *Cell Host Microbe* **10,** 515–526, doi: 10.1016/j.chom.2011.09.012 (2011).
71. Haecker, I. *et al.* Ago HITS-CLIP expands understanding of Kaposi's sarcoma-associated herpesvirus miRNA function in primary effusion lymphomas. *PLoS Pathog* **8,** e1002884, doi: 10.1371/journal.ppat.1002884 (2012).
72. Skalsky, R. L. *et al.* The viral and cellular microRNA targetome in lymphoblastoid cell lines. *PLoS Pathog* **8,** e1002484, doi: 10.1371/journal.ppat.1002484 (2012).
73. Xue, Y. *et al.* Direct conversion of fibroblasts to neurons by reprogramming PTB-regulated microRNA circuits. *Cell* **152,** 82–96, doi: 10.1016/j.cell.2012.11.045 (2013).
74. Karginov, F. V. & Hannon, G. J. Remodeling of Ago2-mRNA interactions upon cellular stress reflects miRNA complementarity and correlates with altered translation rates. *Genes & development* **27,** 1624–1632, doi: 10.1101/gad.215939.113 (2013).
75. Lipchina, I. *et al.* Genome-wide identification of microRNA targets in human ES cells reveals a role for miR-302 in modulating BMP response. *Genes & development* **25,** 2173–2186, doi: 10.1101/gad.17221311 (2011).
76. Farazi, T. A. *et al.* Identification of distinct miRNA target regulation between breast cancer molecular subtypes using AGO2-PAR-CLIP and patient datasets. *Genome biology* **15,** R9, doi: 10.1186/gb-2014-15-1-r9 (2014).
77. Baltz, A. G. *et al.* The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell* **46,** 674–690, doi: 10.1016/j.molcel.2012.05.021 (2012).
78. Yokoshi, M. *et al.* Direct binding of Ataxin-2 to distinct elements in 3′ UTRs promotes mRNA stability and protein expression. *Mol Cell* **55,** 186–198, doi: 10.1016/j.molcel.2014.05.022 (2014).
79. Martin, G., Gruber, A. R., Keller, W. & Zavolan, M. Genome-wide analysis of pre-mRNA 3′ end processing reveals a decisive role of human cleavage factor I in the regulation of 3′ UTR length. *Cell Rep* **1,** 753–763, doi: 10.1016/j.celrep.2012.05.003 (2012).
80. Yao, C. *et al.* Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. *Proceedings of the National Academy of Sciences of the United States of America* **109,** 18773–18778, doi: 10.1073/pnas.1211101109 (2012).
81. Macias, S. *et al.* DGCR8 HITS-CLIP reveals novel functions for the Microprocessor. *Nat Struct Mol Biol* **19,** 760–766, doi: 10.1038/nsmb.2344 (2012).
82. Sauliere, J. *et al.* CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. *Nat Struct Mol Biol* **19,** 1124–1131, doi: 10.1038/nsmb.2420 (2012).
83. Mukherjee, N. *et al.* Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol Cell* **43,** 327–339, doi: 10.1016/j.molcel.2011.06.007 (2011).
84. Lebedeva, S. *et al.* Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Mol Cell* **43,** 340–352, doi: 10.1016/j.molcel.2011.06.008 (2011).
85. Friedersdorf, M. B. & Keene, J. D. Advancing the functional utility of PAR-CLIP by quantifying background binding to mRNAs and lncRNAs. *Genome biology* **15,** R2, doi: 10.1186/gb-2014-15-1-r2 (2014).
86. Paronetto, M. P. *et al.* Regulation of FAS exon definition and apoptosis by the Ewing sarcoma protein. *Cell Rep* **7,** 1211–1226, doi: 10.1016/j.celrep.2014.03.077 (2014).
87. Hoell, J. I. *et al.* RNA targets of wild-type and mutant FET family proteins. *Nat Struct Mol Biol* **18,** 1428–1431, doi: 10.1038/nsmb.2163 (2011).
88. Kaneko, S., Son, J., Shen, S. S., Reinberg, D. & Bonasio, R. PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat Struct Mol Biol* **20,** 1258–1264, doi: 10.1038/nsmb.2700 (2013).
89. Kishore, S. *et al.* Insights into snoRNA biogenesis and processing from PAR-CLIP of snoRNA core proteins and small RNA sequencing. *Genome biology* **14,** R45, doi: 10.1186/gb-2013-14-5-r45 (2013).
90. Ascano, M., Jr. *et al.* FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* **492,** 382–386, doi: 10.1038/nature11737 (2012).
91. Lagier-Tourenne, C. *et al.* Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nat Neurosci* **15,** 1488–1497, doi: 10.1038/nn.3230 (2012).
92. Nakaya, T., Alexiou, P., Maragkakis, M., Chang, A. & Mourelatos, Z. FUS regulates genes coding for RNA-binding proteins in neurons by binding to their highly conserved introns. *Rna* **19,** 498–509, doi: 10.1261/rna.037804.112 (2013).
93. Huelga, S. C. *et al.* Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep* **1,** 167–178, doi: 10.1016/j.celrep.2012.02.001 (2012).

94. Zarnack, K. *et al.* Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* **152,** 453–466, doi: 10.1016/j.cell.2012.12.023 (2013).
95. Konig, J. *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* **17,** 909–915, doi: 10.1038/nsmb.1838 (2010).
96. Yoon, J. H. *et al.* PAR-CLIP analysis uncovers AUF1 impact on target RNA fate and genome integrity. *Nat Commun* **5,** 5248, doi: 10.1038/ncomms6248 (2014).
97. Katz, Y., Wang, E. T., Airoldi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* **7,** 1009–1015, doi: 10.1038/nmeth.1528 (2010).
98. Xiao, R. *et al.* Nuclear matrix factor hnRNP U/SAF-A exerts a global control of alternative splicing by regulating U2 snRNP maturation. *Mol Cell* **45,** 656–668, doi: 10.1016/j.molcel.2012.01.009 (2012).
99. Hafner, M. *et al.* Identification of mRNAs bound and regulated by human LIN28 proteins and molecular requirements for RNA recognition. *Rna* **19,** 613–626, doi: 10.1261/rna.036491.112 (2013).
100. Graf, R. *et al.* Identification of LIN28B-bound mRNAs reveals features of target recognition and regulation. *RNA Biol* **10,** 1146–1159, doi: 10.4161/rna.25194 (2013).
101. Sievers, C., Schlumpf, T., Sawarkar, R., Comoglio, F. & Paro, R. Mixture models and wavelet transforms reveal high confidence RNA-protein interaction sites in MOV10 PAR-CLIP data. *Nucleic acids research* **40,** e160, doi: 10.1093/nar/gks697 (2012).
102. Xue, Y. *et al.* Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol Cell* **36,** 996–1006, doi: 10.1016/j.molcel.2009.12.003 (2009).
103. Raj, B. *et al.* A global regulatory mechanism for activating an exon network required for neurogenesis. *Mol Cell* **56,** 90–103, doi: 10.1016/j.molcel.2014.08.011 (2014).
104. Ibrahim, F. *et al.* Identification of *in vivo*, conserved, TAF15 RNA binding sites reveals the impact of TAF15 on the neuronal transcriptome. *Cell Rep* **3,** 301–308, doi: 10.1016/j.celrep.2013.01.021 (2013).
105. Tollervey, J. R. *et al.* Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat Neurosci* **14,** 452–458, doi: 10.1038/nn.2778 (2011).
106. Wang, Z. *et al.* iCLIP predicts the dual splicing effects of TIA-RNA interactions. *PLoS biology* **8,** e1000530, doi: 10.1371/journal.pbio.1000530 (2010).
107. Schonemann, L. *et al.* Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes & development* **28,** 2381–2393, doi: 10.1101/gad.250985.114 (2014).
108. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505,** 117–120, doi: 10.1038/nature12730 (2014).

Acknowledgements

The authors wish to thank the members of Janga lab for their comments and feedback on a previous version of this manuscript. SCJ acknowledges support from the School of Informatics and Computing at IUPUI.

Author Contributions

S.N. carried out the computational analysis related to human data and wrote different parts of the manuscript. Y.N. carried out the computational analysis related to the yeast genome. S.C.J. supervised and designed research, analyzed data and wrote parts of the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Nishtala, S. *et al.* Dissecting the expression relationships between RNA-binding proteins and their cognate targets in eukaryotic post-transcriptional regulatory networks. *Sci. Rep.* **6**, 25711; doi: 10.1038/srep25711 (2016).